Territorial grids in the Mediterranean: space versus population

Yoann Doignon, Sébastien Oliveau

EUT EDIZIONI UNIVERSITÀ DI TRIESTE

Bollettino della ASSOCIAZIONE ITALIANA di CARTOGRAFIA 2015 (154), 46-63

ISSN 2282-472X (online)
ISSN 0044-9733 (print)
http://www.openstarts.units.it/dspace/handle/10077/9933

DOI: 10.13137/2282-472X/11827

Reti territoriali nel Mediterraneo: territorio vs. popolazione*

*Territorial grids in the Mediterranean: space versus population**

YOANN DOIGNON*, SEBASTIEN OLIVEAU**

* Université Nice Sophia Antipolis, Avignon Université, Aix Marseille Université, CNRS, ESPACE UMR 7300 – LAMES UMR 7305 – DEMOMED – yoann.doignon@univ-amu.fr

** Université Nice Sophia Antipolis, Avignon Université, Aix Marseille Université, CNRS, ESPACE UMR 7300 - DEMOMED ebastien.oliveau@univ-amu.fr

DEMOMED: Observatoire Démographique de la Méditerranée – www.demomed.org

Riassunto

Questo articolo esamina come fare confronti internazionali su scala mediterranea in una prospettiva demografica e geografica. Questo studio si concentra sulla variabilità osservata delle superfici e delle popolazioni di diverse reti disponibili per i paesi mediterranei.
Ci chiediamo in primo luogo su ciò che rende rilevante una griglia rispetto ad un'altra. Inoltre ci proponiamo di utilizzare l'autocorrelazione spaziale come indicatore di qualità delle caratteristiche statistiche e spaziali di una griglia per una data variabile. Basandoci su un confronto tra diverse griglia infra-nazionali, tra cui due reti realizzate ad hoc, esploriamo la variabilità generata dalla griglia prescelta su una variabile che unisce lo spazio e la popolazione: la densità. I risultati sottolineano l'importanza di questo tipo di approccio prima di qualsiasi opera di confronto internazionale e ricorda la potenziale importanza che può assumere il MAUP (Modifiable Areal Unit Problem) negli studi spaziali.

Parole chiave

Densità, reti territoriali, Mediterraneo, autocorrelazione spaziale, MAUP, cartografia, SIG

Abstract

This paper explains how to make international comparisons on a Mediterranean scale for demographic and geographical themes. This study deals with the variability observed of surfaces and populations of the grids available of the Mediterranean countries.
First, we wonder what makes the relevance of a grid compared to another. Then, we use spatial autocorrelation as an indicator of spatial and statistical characteristics of a grid for a specific variable. We create two grids for the occasion. Comparing several grids, we explore the variability generated by the variable of population density. The results show the importance of this kind of approach before beginning any international comparison. They also remind the importance of MAUP in spatial studies.

Keywords

Density, territorial grid, the Mediterranean, spatial autocorrelation, MAUP, cartography, GIS

* This work is part of SoDeMoMed project, financed by ANR Transmed (ANR-12-TMED-0005).
The authors are grateful to Elena Ambrosetti for her Italian translation.

1. Introduction

While data of all types is becoming increasingly available and although it is usually geo-referenced and the tools are easier to manage, there is often little reflection on the way it is used. This is due to a lack of user training or simply negligence. Among this data, that relating to population is more easily accessible (census, surveys) and is often linked to suitable map collections.

For this reason, it is tempting to study the spatial dimensions of the phenomena under consideration and this extends across all disciplines and beyond academia. Mapping is made accessible to everyone. Even though most geographers and cartographers are aware of the pitfalls of mapping and the importance that should be placed on prior analysis of the variables (statistical study and discretisation), they very rarely focus on the grids that constitute the framework of their study.

This means that the necessary deconstruction of the grids is carried out all too rarely. People do not question often enough the consequences of the default choices made. We will examine how, in addition to an academic exercise, this is also a broader reflection on the 'proper' level of analysis which is then presented to us. Secondly, we will examine possible alternatives for determining this proper level. Finally, we will apply our remarks and recommendations to a pan-Mediterranean analysis of population densities in order to provide a concrete demonstration of the effect on perceptions and measurements of the phenomenon of the grids adopted.

2. What is a proper level of analysis?

Until recently, the availability of data limited the possible levels of observation to a great extent. However, wider dissemination of data means this is becoming less the case. Whereas the lack of data previously left the researcher with little choice about the level of study, the diversity of scales and levels of observations that are now available mean the researcher is obliged to make a selection. Although having a choice of levels of representation may look like a comfortable position, it raises new methodological problems. At the forefront of these is what most geographers have been calling the Modifiable Areal Unit Problem (MAUP) for 40 years.

2.1 The MAUP and its implications in a geographic search

The MAUP is a recurring theme in spatial analysis. Stan Openshaw (and Peter Taylor) devised the term MAUP in their 1979 article: "a million or so correlation coefficients: three experiments on the Modifiable Areal Unit Problem." The idea, later developed in more detail (Openshaw, 1984), is that statistical measurements of spaces are directly dependent on the resolution and shape of the grids used. These two effects are therefore usually differentiated by using the terms scale on one hand, and zoning effect on the other.

The scale effect emphasizes the variations experienced by data when the level of observation is altered, specifically, the statistical smoothing effect that occurs when data is aggregated. The disappearance of detail as it "ascends" the geographic levels and its contribution to an understanding of the phenomena and their structures have already been well highlighted. Arbia, for example, when he alludes to the possibility of a second law of geography, suggests that: "Everything is related to everything else, but things observed at a coarse spatial resolution are more related than things observed at a finer resolution." (Arbia et al., 1996)[1].

The zoning effect highlights the effect of forms of territorial breakdown on the results. Thus, during the aggregation process, the number of possibilities for linking the grids and forming new ones and the fact that a large number of grids exists assume greater importance. The number of possible combinations increases very rapidly and each solution produces different values for the newly obtained grids. The zoning effect characterizes administrative divisions (particularly electoral divisions) and is added to the scale effect.

The MAUP affects measurements and thus it has implications for statistics. Thus, Gehlke & Bielh (1934, p. 170), without referring to the MAUP, already noted

1 As a reminder, the first law of geography formulated by Tobler is: "Everything is related to everything else, but near things are more related than distant things." (Tobler, 1970).

ISSN 2282-472X (online)
ISSN 0044-9733 (print)

that "variations in the size of the correlation coefficient seem conditioned upon changes in the size of the unit used, with a smaller value of *r* associated with the smallest unit rather than with the largest. Various ways of grouping have considerable influence on the *r*, as well as has the size of the area."

Even if the issue is a longstanding one, we can see that the problem remains topical. The zoning effect remains little studied and should certainly be examined in greater detail. The scale effect is easier to measure but it also merits a specific study in order to take into account its effect on measurements (Reynolds, 1998).

There is also another issue: discovering whether the MAUP, rather than constituting a 'problem', is actually a constituent of spatial data and, as such, an item of information rather than a problem (Grasland & Madelin, 2006). Indeed, once one is aware of the existence of this phenomenon of variability as a result of aggregation, exploring it can represent a new source of information.

Behind these questions lies a more general question: whether on any given scale there is a level of analysis which is better than the others. Before returning to what we might mean by "better", we should recall that this type of question has already been explored.

When seeking an optimal grid, Openshaw (1977) proposed a method of automatic partition to detect which spatial sets are suitable for a particular purpose. He showed that the optimal zoning for one variable is not necessarily the same for another, and that the results for a study variable implicitly depend on the level and scale at which they are measured. Significance testing procedures have been developed to ensure the robustness of results during changes in levels and aggregation (Knudsen, 1987). Getis spatial statistics are used to detect any effects of the MAUP on a set of socio-economic variables (Amrhein, Reynolds, 1996). Various methods are used which specifically address the scale effect: fractal analysis and geo-statistical methods such as kriging, or variograms (Tate and Atkinson, 2001), GWR (Geographically Weighted Regression) to account for spatial heterogeneity (Fotheringham and al., 2000), or even a method that reduces the scale effect to a minimum by searching for strong internal homogeneity (Holt, Steel, Tranmer, 1996).

2.2 A 'better' level of observation?

The following reflections depart from the premise that there is a level of observation for each variable that is better than the others. But, what exactly is meant by better? Here again, one can adopt a number of positions, and we will see that, in terms of the socio-demographic data that interest us here, there are different points of view. Firstly, the wide distribution of micro level data (such as census districts) has led many researchers to assume that the most precise data is the best. Because it offers more detail it is therefore more valuable than all the others. Nevertheless, we soon become aware that, for some variables, too fine a grid adds statistical noise that is related to the low number of individuals per grid square and it therefore tends to present a blurred vision of the territories being studied (for a demographic example, see Guilmoto, Oliveau, 2007). Then there is another position, presented in detail below, which involves choosing the level that presents the most readable spatial structures. This ad hoc approach makes it easier to put forward geographical explanations and it takes a largely empirical approach, which also necessarily raises a number of questions.

Beyond this general questioning, another issue arises when one seeks to compare different territories for a specific level of observation. Therefore, as this paper proposes, this may make international comparisons difficult. The right level for one country will not necessarily be the same for several countries.

The existing levels for observing one territory are not necessarily the same for a second territory. It then becomes necessary to suggest equivalents which are not precise and which result in new mechanisms for selecting the 'best' level. Indeed, one could choose a set of grids in order to reduce the measured statistical variability or attempt to obtain a regular grid based on its surface area. The first approach corresponds to an aspatial approach, while the second disregards the statistical dimension. However, we propose an alternative based on measuring the spatial autocorrelation of the studied variables. This takes into account the statistical variability of the variable based on the spatial distribution of the grids.

ISSN 2282-472X (online)
ISSN 0044-9733 (print)

2.3 Measuring spatial autocorrelation as an exploratory tool for determining the best analysis levels

The purpose of this article is not to review the methods that allow us to measure spatial autocorrelation. Extensive literature on the subject has been available since the publication of the classic work by Cliff and Ord in 1973. However, we would like to stress that we have based our measurements on Moran's index (Moran's *I*), used in the form suggested by Cliff & Ord (1981). This is the one found in most publications and it is more widely than that published by Moran (1950).

Thus, Moran's I measures the covariation of the values of a point and that of its neighbours by returning the result to the variance of the set of points. The result of Moran's I calculation is easy to interpret as it is interpreted approximately as a classical linear correlation coefficient. It varies from -1 (negative spatial autocorrelation: the values of the variables for individual neighbours are set against the average) and +1 (positive spatial autocorrelation: the neighbours are similar). It should however be noted that Moran's I value can sometimes be greater than 1 or less than -1. Thus it is not strictly restricted limited to -1 and + 1. Zero[2] marks the absence of negative or positive spatial autocorrelation (at least on a global scale).

Moran's index

Insert: Moran's I

$$\textit{Moran's}\ \mathrm{I}\ \frac{\mathrm{cov}}{\mathrm{var}} = \mathrm{let}\ \mathrm{cov} \div \mathrm{var}$$

$$\textit{Moran's}\ \mathrm{I} = \frac{\sum_i \sum_j w_{ij}(z_i - \bar{z})(z_j - \bar{z})}{m} \div \frac{\sum_i (z_i - \bar{z})^2}{n}$$

$$\textit{Moran's}\ \mathrm{I} = \frac{\sum_i \sum_j w_{ij}(z_i - \bar{z})(z_j - \bar{z})}{m} \times \frac{n}{\sum_i (z_i - \bar{z})^2}$$

$$\textit{Moran's}\ \mathrm{I} = \frac{n}{m} \times \frac{\sum_i \sum_j w_{ij}(z_i - \bar{z})(z_j - \bar{z})}{\sum_i (z_i - \bar{z})^2}$$

We are in a field of spatial statistics and the letters "x" and "y" are reserved for the individual coordinates.

z_i = value of the variable at point "i" and mean $\bar{z}$

i = individual

j = neighbours of "i" individuals.

n = total number of individuals in the sample

m = total number of pairs of neighbours

W = weighting matrix, whose elements are, for example, set to "1" for "i, j" neighbours and otherwise "0".

Moran's *I* makes it possible to measure the spatial structure of a socio-demographic phenomenon. In other words, to reveal whether the studied phenomenon reveals a particular distribution in space or not, and to what degree. Highlighting a spatial structure is therefore the starting point for the geographic analysis, which aims to explain it.

By inverting the reasoning, we propose to define the best level of analysis as the one showing the highest level of spatial autocorrelation. One may well consider that the phenomenon is best revealed at the level presenting the most marked spatial structure. Based on this assumption, we examined the spatial structure of density in France on several dates (Oliveau et al. 2013). This produced enough evidence on the levels that are most likely to demonstrate the phenomena in their spatial dimension.

The following results (see Figure 1) show that the most suitable level for studying densities in France, regardless of the period, would be the district level.

Based on these results, we decided to use the spatial autocorrelation measurement as an indicator on the Mediterranean scale to assess the relevance of the grids that are available for different countries.

3. Creating a sub-national grid around the Mediterranean

In this section, we look at different ways of creating a sub-national grid in the Mediterranean, understood here

2 It will be recalled that, in fact, it is not set to 0 but -1/(n-1), which therefore moves very quickly to 0.

ISSN 2282-472X (online)
ISSN 0044-9733 (print)

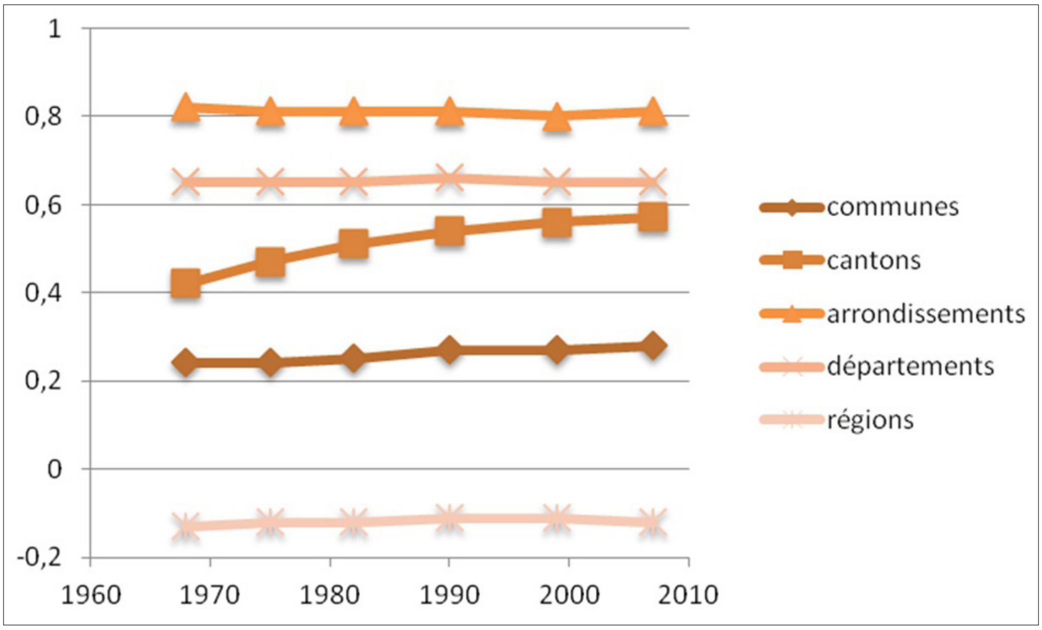

FIGURE 1
Moran's I according to the administrative level in France for population density

AUTHORS: Doignon Y., Oliveau S.

SOURCES: Insee (1968, 1975, 1982, 1990, 1999, 2009)

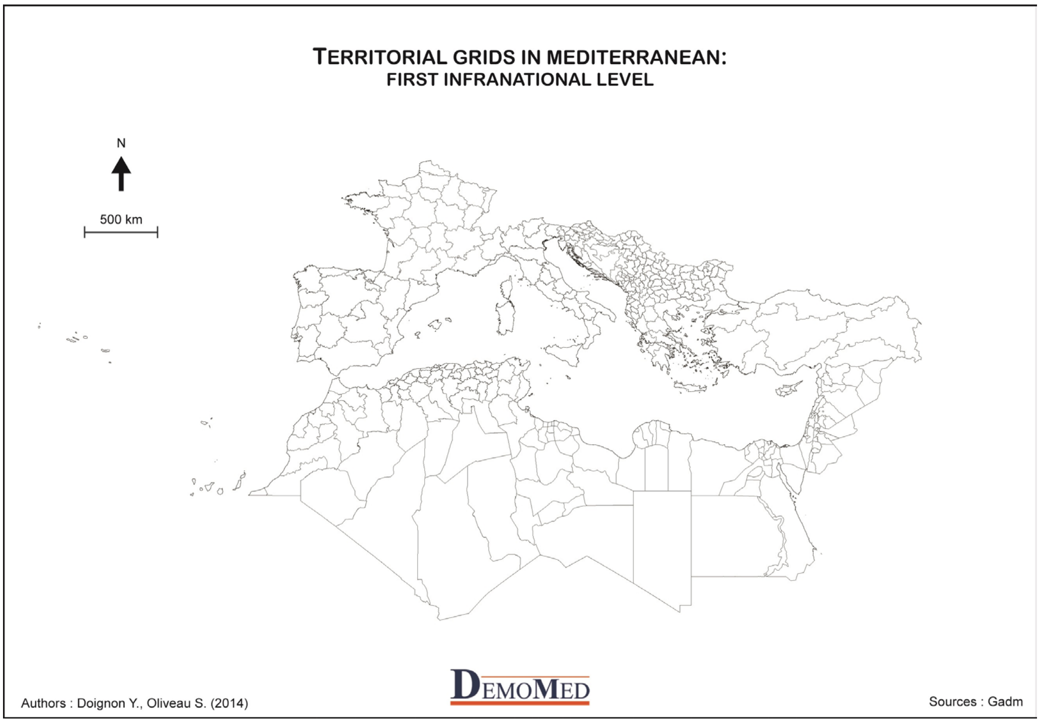

FIGURE 2
Grid of the first administrative level in the Mediterranean

AUTHORS: Doignon Y., Oliveau S.

SOURCES: Gadm v2, 2014

ISSN 2282-472X (online)
ISSN 0044-9733 (print)

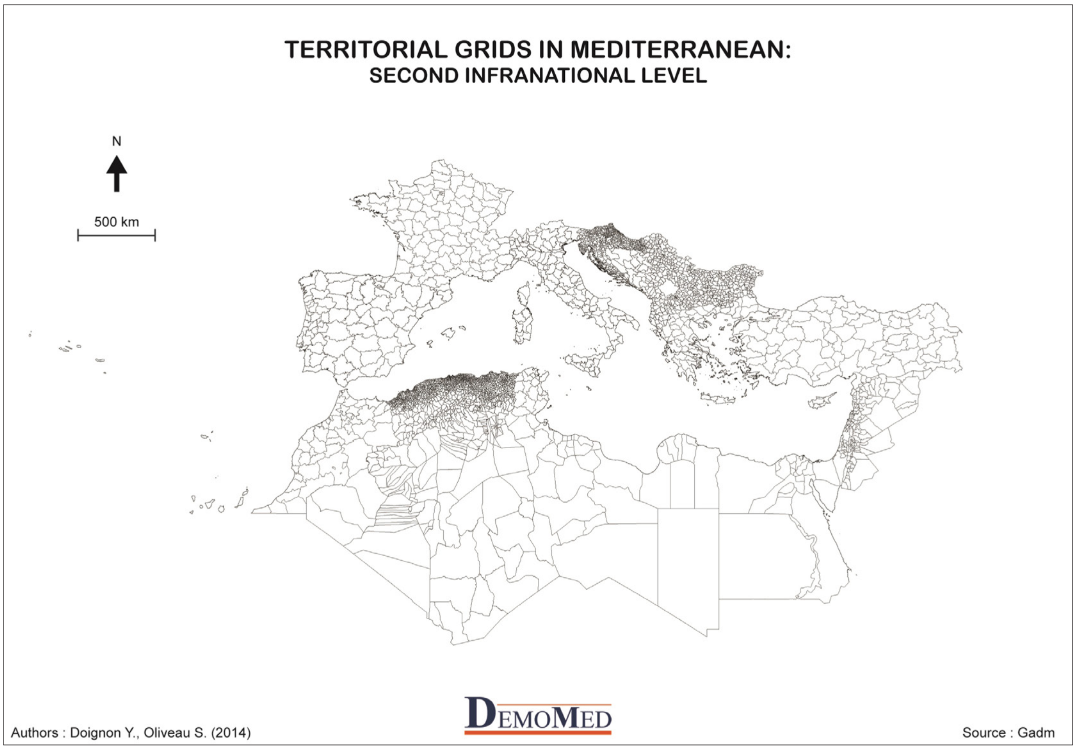

FIGURE 3
Grid of the second administrative level in the Mediterranean

AUTHORS: Doignon Y., Oliveau S.

SOURCES: Gadm v2, 2014

as a group of 27 countries[3]. Specifically, we need to select an administrative level in each country. The number of possible combinations is therefore extremely important. We propose creating several grids and we will assess the relevance of each in Section 3.

Firstly, we present a standard approach that consists of selecting the same administrative level in each country, by analogy to the European NUTS. Secondly, we outline a method for the harmonization of the grids based on a specific criterion.

The data used in this article is detailed in Appendix 1.

3.1 Standard approach: selecting the same administrative level in each country

This approach involves selecting the same administrative level in each country. For example, we select the first administrative level, i.e. the level below the State. In France, Spain, Italy and Morocco this is the region, in Algeria, the *wilaya*, and the *mohafazat* in Egypt.

The assumption underlying this method is the equivalence of the countries' administrative levels based on their ranking. The Italian regions correspond to the Albanian *qark* because this is the first administrative level in each of these countries.

In the case of the Mediterranean, Figures 2 and 3 show the first and the second administrative levels in all the Mediterranean countries, respectively. The extreme heterogeneity of the size of the grids is striking. The Spanish provinces have an average surface area which corresponds to that in Lebanon (see table 5).

This shows us that the hypothesis of the equivalence of the grids according to the order of the administrative levels is not realistic within the context of their international comparability. It is therefore necessary to perform harmonization of the selected grids. Equivalence between the various administrative levels in Mediterranean countries needs to be defined. In order to compare the Italian provinces, which administrative level should

3 Mediterranean refers here to all the following countries: Morocco, Algeria, Tunisia, Libya, Egypt, Lebanon, Israel, Syria, Palestine, Jordan, Turkey, Cyprus, Greece, Albania, Macedonia, Kosovo, Montenegro, Bulgaria, Serbia, Bosnia and Herzegovina, Croatia, Slovenia, Italy, Malta, France, Spain, Portugal.

ISSN 2282-472X (online)
ISSN 0044-9733 (print)

one choose in Lebanon or Libya in order to obtain comparable entities? The criteria used to determine the "comparability" of the grids also need to be questioned.

3.2 An alternative approach: harmonizing grids based on a common criterion

Simply adjusting the order of the administrative levels in order to construct a sub-national grid in the Mediterranean is not enough. It is necessary to establish equivalence between the grids in order to select comparable administrative levels in each country. This harmonization needs to put forward a criterion that defines different grids as being equivalent or at least comparable. Two criteria spring to mind: surface area and population. The comparability criterion is very dependent on the research problematic. In geography, we tend to rely heavily on surface area, whereas demographers and sociologists are more likely to use population.

In order to compare grids in different countries, we propose calculating the elements of centrality (mean and median) and dispersion (coefficient of variation) of the surface area and population for each administrative level. Tables 4 and 5 (in the appendices) confirm the extreme heterogeneity previously observed in terms of the surface area of the selected grids (Figures 2 and 3). The first administrative levels are not equivalent in terms of population or surface area. Whereas the Italian regions average 15,000 km² with an average of 2,990,000 inhabitants, Serbian *okrug* average 3,100 km² with 300,000 inhabitants. In this case, we should, for example, select those Italian provinces with an average size of 2,800 km² and an average of 566,000 inhabitants. This means that, in order to obtain broadly comparable grids for these two countries, one would need to select the second administrative level in Italy and the first level in Serbia.

These tables highlight another important aspect, namely the internal heterogeneity of some grids. The dispersion indicator can be very pronounced in some countries. This is the case of the Algerian *wilayas*, for example, which have a surface area coefficient of variation of 2.39. This situation is understandable given that the Saharan *wilayas* are large whereas those on the Mediterranean coast are small. In this case, the mean does not really make sense and greater importance should be attached to the median. The researcher should use both indicators to assess a grid.

These grids showing marked heterogeneity on a national scale are problematic in that would be more accurate to create a grid consistent with the administrative levels. One solution would be to homogenize the grid by spatial aggregation and disaggregation of its units (as was done for India and China, Guilmoto and Oliveau, 2007). This method significantly increases the internal homogeneity of the grids. However, the units obtained by this process can be meaningless in situations where the initial administrative divisions have a precise, known legal meaning (Dumolard, 1998). This can prove particularly problematic when it comes to interpreting results.

Another solution is to accommodate the existing heterogeneity in favour of a higher interpretative potential when the analysis is complete. On the other hand, it should be remembered that the grids in some countries are heterogeneous and this should be taken into account when interpreting the data.

This type of summary table facilitates the creation of a similar grid based on a specific criterion. It is also possible to harmonize using a subsequently defined value. In this paper, we construct two grids of this type: the first is harmonized to 500,000 inhabitants, a second, harmonized to around 25,000 km². In each country, we retain the administrative level that is closest to these values in terms of the mean and median of the grids. The administrative levels for our two grids are recorded in Table 1. For some countries, all sub-national grids have a population or a surface area that is very different to the criterion value. Accordingly, the administrative level selected is the state level.

In concrete terms, the grids created here are not perfectly harmonized because not all the administrative levels selected have the expected population or surface area. For the Mediterranean, this is due to the impossibility of comparing grids in some countries, regardless of the administrative level under consideration. We illustrate this aspect with two examples. If we attempt to compare Algeria and Spain in terms of population, we can see in Table 4 that the first Spanish level is very densely populated compared with the first Alge-

ISSN 2282-472X (online)
ISSN 0044-9733 (print)

rian level. On the other hand, the second level corresponds much better: a median of almost 550,000 compared with 565,000 in Algeria. Thus, these two grids are comparable in terms of population.

However, if we want to compare Algeria and Portugal in terms of surface area, Appendix 1 shows that it is impossible to find a suitable match. Whereas the first Algerian level is 3.5 times larger than the Portuguese one, the second level is twice as small as the Portuguese second level and almost nine times the size of the first level. In concrete terms, we are unable to find comparable grids in Algeria and Portugal. Having discarded the idea of modifying the administrative grid, we need to reach a compromise, bearing in mind that the comparison between these two countries is not optimal.

Finally, in this article we have four levels of grid available for the Mediterranean: a grid with the first administrative level in all countries; a grid with the second administrative level in all countries for which population data is available (if it is not, the higher level is selected); a grid harmonized to 500,000 inhabitants and finally a grid harmonized to 25,000 km². Table 1 summarizes the administrative levels selected for each of these four grids.

As we have shown, by maintaining a simple approach based on two administrative levels, it is not difficult to produce four different pan-Mediterranean grids. Making a selection will not be easy, however. In addition, bearing in mind what we recalled about the MAUP, it is highly probable that these grids will not have the same properties or the same influence on the results of the analysis. *A priori*, some will be more suited to cartographic representation and others to statistical analysis. We propose to test this hypothesis in the last section.

4. What is the impact of the grids on analysis?

The spatial and statistical variability of these four grids will be studied. We will examine which grid is more homogeneous in terms of its population and surface area. We will also introduce the population density variable in order to observe the effect of each grid on a third variable. For each grid, we calculate the mean, the coefficient of variation and the median of the following variables: population, surface area, and population density. Tables 2 and 3 highlight the large statistical variability of the results according to the grids.

For the population criterion, Grid 2 is the one that shows the greatest dispersion. This is not surprising as a larger scale reveals the higher heterogeneity. The most homogeneous grids are clearly the two harmonized grids. However, these are not situated at the same level of observation. Grid 3 has a median of 360,000 inhabitants while that in Grid 4 is 760,000 inhabitants. In other words, Grid 4 is twice as aggregated as Grid 3. Both are statistically more homogenous but do not present the same level of observation.

For the surface area criterion, Grid 2 remains the one with the most dispersion. Note that this grid does not relate to all second administrative levels in the country, because we selected only the levels for which population data was available. To examine what the dispersal of the area with the second administrative level in all countries would have been, we created Grid 5. In this case, the spread is even greater... Logically, the grid that is harmonized according to surface area (Grid 4) is the one that presents the lowest dispersion.

In the case of population density, the harmonized grids (nos. 3 and 4) are statistically less dispersed. It should be noted that Grid 4 has a median that is comparable to Grid 2, but with a much lower dispersion. The average and median population density is stronger in Grids 1 and 3. One can still observe differences of almost 90 inhabitants/km² depending on the grid, which is an important aspect when one is attempting to examine population distribution.

These figures show that investing in a harmonized grid, even an imperfect one, is provides better value than economizing by using a grid made up of a set of grids that are equivalent from an administrative point of view. Nevertheless, one should remember that a grid which is suitable for statistical analysis is not necessarily the most suitable grid for cartographic representation.

Figures 4, 5, 6 and 7 represent the population density for each of the previously selected grids (except Grid 5, for which we do not have all the data). They are a very good example of the scale effect of the MAUP: depending on the map selected, the interpretations of popula-

ISSN 2282-472X (online)
ISSN 0044-9733 (print)

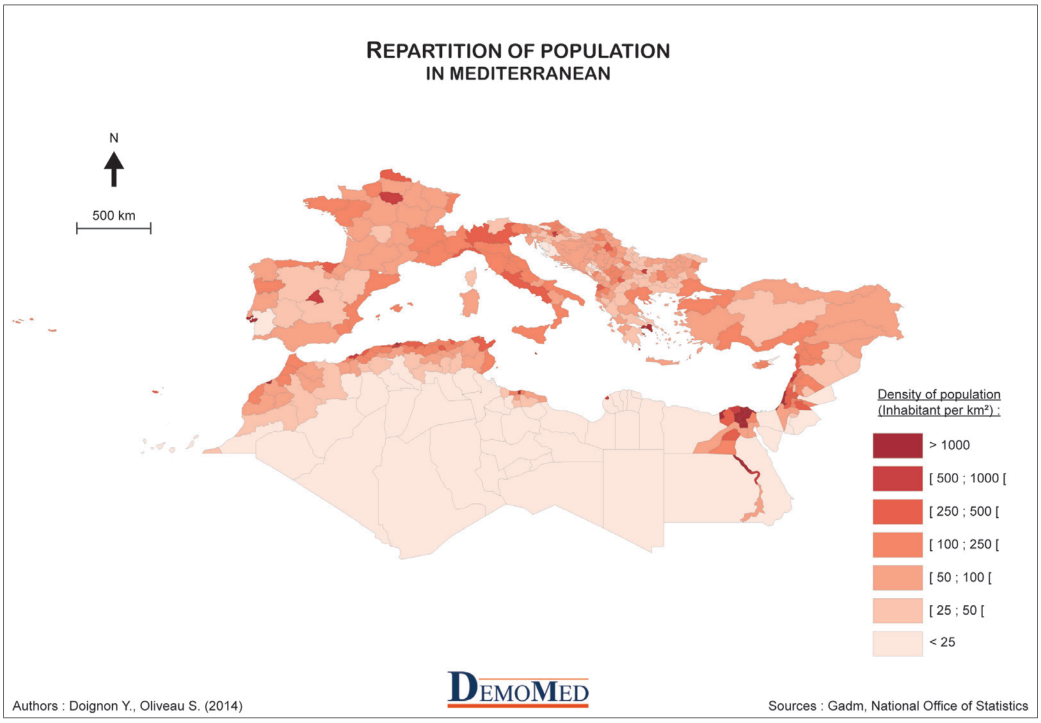

FIGURE 4
Population density with Grid 1

AUTHORS: Doignon Y., Oliveau S.

SOURCES: Gadm v2, most recent estimates from National Office of statistics

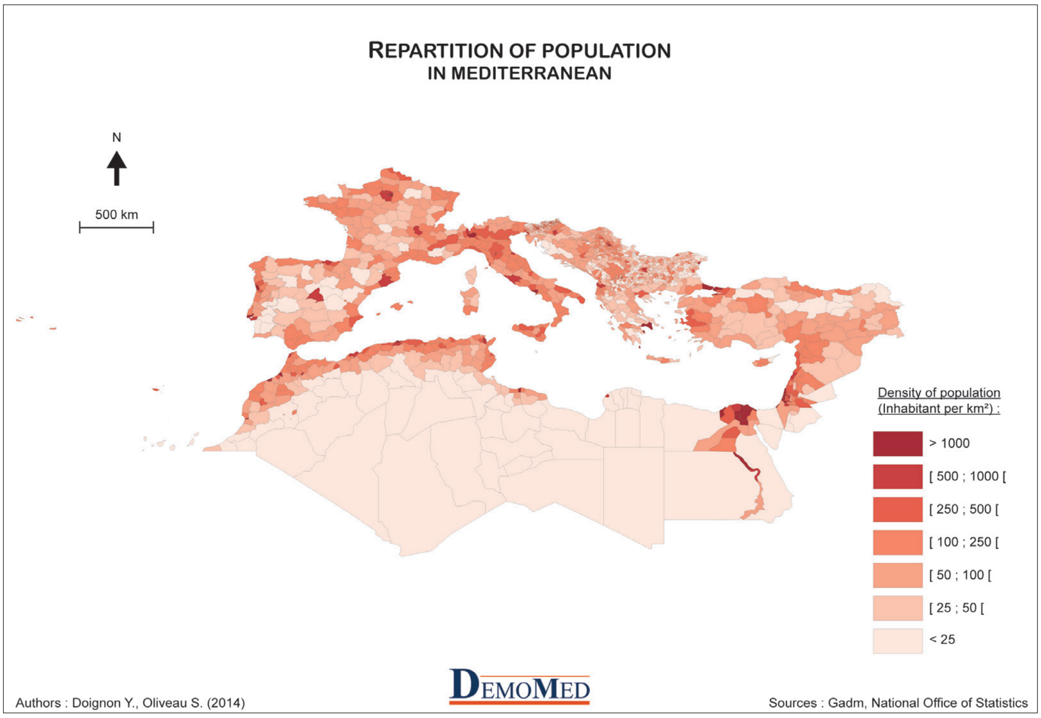

FIGURE 5
Population density with Grid 2

AUTHORS: Doignon Y., Oliveau S.

SOURCES: Gadm v2, most recent estimates from National Office of statistics

ISSN 2282-472X (online)
ISSN 0044-9733 (print)

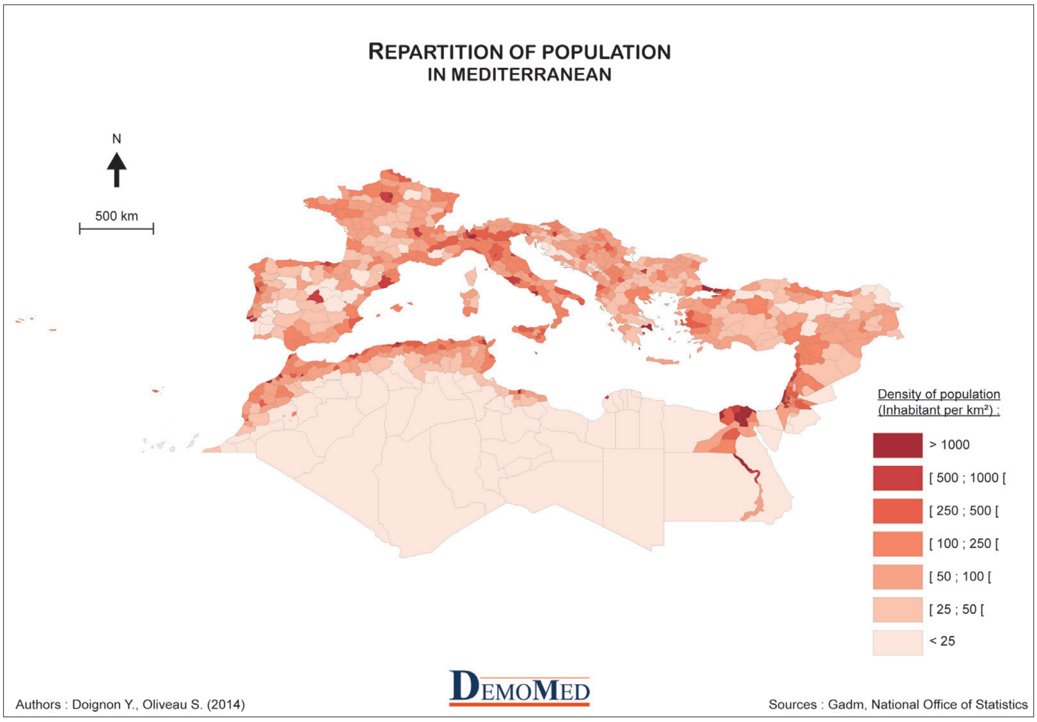

FIGURE 6
Population density with Grid 3

AUTHORS: Doignon Y., Oliveau S.

SOURCES: Gadm v2, most recent estimates from National Office of statistics

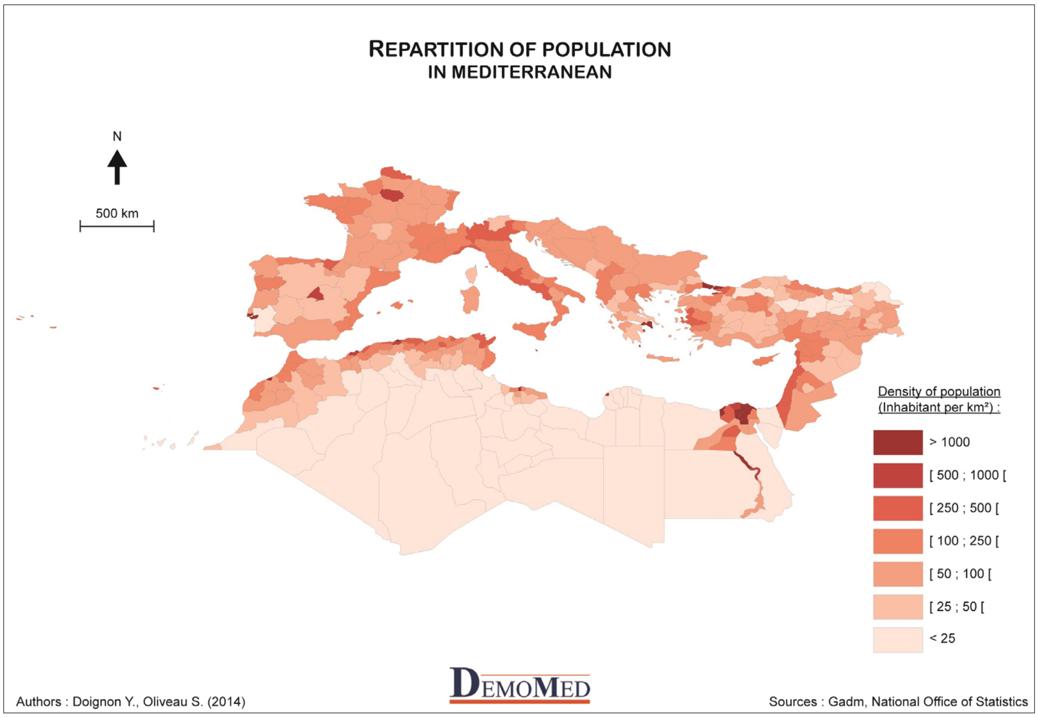

FIGURE 7
Population density with Grid 4

AUTHORS: Doignon Y., Oliveau S.

SOURCES: Gadm v2, most recent estimates from National Office of statistics

ISSN 2282-472X (online)
ISSN 0044-9733 (print)

tion distribution vary. In the case of Grid 1, the highest population densities are noted in the regional unit comprising the capitals of European countries (Portugal, Spain, France, Italy, Slovenia and Greece). Furthermore, very few low densities are noted on the north shore. In the case of the south shore, people are located near water: rivers (such as the Nile), coastal areas in North Africa and the Near East. Grid 2 shows more nuanced distribution. In Europe, the diagonal of French empty space and the Spanish desert are revealed, as well as low density areas in Turkey. Several urban centres can be discerned in Morocco. The mapping changes on the south shore in Grid 1 are somewhat limited because, in several countries (Algeria, Libya, Egypt, Syria, Israel and Jordan) we have been obliged to retain the first administrative level due to lack of data. Therefore the observations are the same for both grids. The extreme dispersion previously observed in terms of the size of units in Grid 2 is easily observed on account of the mapping. In the Balkans, for example, the grids are much more detailed in terms of surface area than in southern Europe or the Maghreb where the grids look much more homogeneous.

Grid no. 3 appears more homogeneous. However, the grids in some regions are less detailed than other grids. Population distribution in the Balkans for example, is made up of higher concentrations in Albania and Kosovo and in Athens, Greece. In the rest of the Balkans, the densities are relatively homogeneous.

Grid 4, meanwhile, offers a compromise between Grids 1 and 2. We can observe the concentrations of population in metropolitan areas and near water (rivers and coastline), but also in unpopulated areas (France, Spain, Turkey, and the Sahara).

The grid used to study population distribution in the Mediterranean directly affects the results of cartographic analysis and map interpretation. A map will be even more readable and interpretable if it contains pronounced spatial structures. The distribution of males in the Maghreb is easily to interpret due to the pronounced spatial structure, i.e. the coast/Sahara contrast. The Balkans present a contrasting example in the case of Grid 2, in which few spatial structures seem to emerge.

However, it is more interesting to measure these spatial structures instead of trying to read them off a choropleth map. To do this, we will build on Moran's I using a neighbourhood by adjacency matrix (see Table 3 and Figure 8). The results are less marked than the previous ones, but it is still possible to determine trends. Both harmonized grids present the highest spatial autocorrelation, which is consistent with the observations made earlier. For order 1 adjacency, Moran's I of Grid 3 reaches 0.56. This result is even more significant as it covers more than 750 units. In addition, the correlogram (see Figure 8) shows that Grid 3 has a sharper spatial structure than the other grids, regardless of the order of adjacency retained. The second harmonized

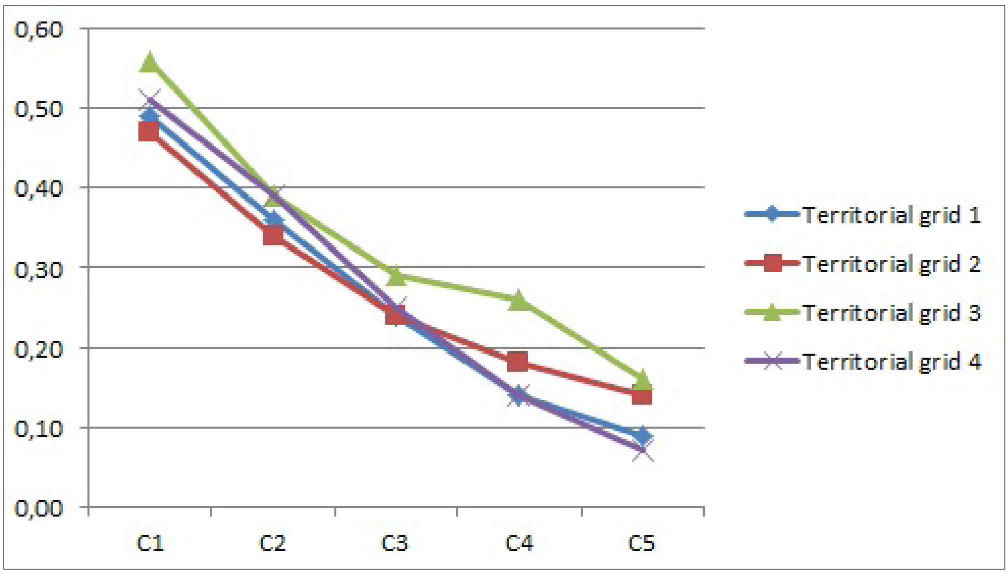

FIGURE 8
Correlogram indices of population density (order of adjacency) Moran

AUTHORS: Doignon Y., Oliveau S.

SOURCES: Gadm v2, most recent estimates from National Office of statistics

ISSN 2282-472X (online)
ISSN 0044-9733 (print)

grid (Grid 4), also shows a stronger spatial autocorrelation than the grids that are not harmonized, but this is less significant than in Grid 3, even though it is more aggregated and therefore more smoothed.

Without being able to demonstrate this, one can consider that the lower Moran's I in Grid 2 is partly the result of the heterogeneity of the grids used (where level 2 is not generally available).

These few succinct procedures show how important the choice of grid is for a geographical study. Indeed, for a similar study area and variable, the grids do not have the same degree of spatial structure. This is a not insignificant aspect when you consider the potential impact on cartographic interpretation (visual over-evaluation of specific spaces, "omitting" areas that are too small). Unsurprisingly, the harmonized grids (and in preference, No. 3) appear to be the most appropriate ones.

5. Conclusion

Although it is frequently overlooked, the choice of grid in statistical or cartographic studies is essential. It will inevitably have an effect on the results. Research into the MAUP has unsuccessfully attempted to minimize its effects. More recently, researchers have begun to accept the pitfalls inherent in spatial data and to propose solutions aimed at taking into account the MAUP as it is not possible to dispense with it. Therefore, the aim is not to disregard the MAUP, but to be aware of its existence in order to seek solutions that limit it or at least integrate it into considerations.

However, an awareness of the bias represented by space in statistical analysis does not justify a failure to question the grid prior to studying it. This reflex questioning of the source should be developed, regardless of whether one is working on one or several countries. However, it becomes even more important when the study focuses on a set of countries at the subnational level. The number of possible grids is then multiplied and the solutions quickly gain in complexity as each country has several different administrative levels. The potential number of combinations for creating an international grid at the subnational level can very soon prove to be huge.

Nevertheless, there are solutions based, in the first instance, on an expert approach - by initially eliminating the levels that are unsatisfactory and, secondly, selecting the most relevant levels based on statistical and spatial homogeneity.

In this article, we have attempted to highlight the need for studies on grids and also to propose realistic solutions based on geo-statistical methods which are easily accessible nowadays. We have used data constructed for the Mediterranean region in order to highlight the extent to which the grid influences the spatial and statistical variability of results. This study has also showed that, when developing rigorous analyses from a statistical and spatial point of view, grids that are 'harmonized' to population or surface area are more effective than grids constructed according to the administrative level.

ISSN 2282-472X (online)
ISSN 0044-9733 (print)

Tables

TABLE 1 – Composition of territorial grids used in this article

	Subnational level				
	Territorial grid 1	**Territorial grid 2**	**Territorial grid 3**	**Territorial grid 4**	**Territorial grid 5**
Albania	1	2	1	0	2
Algeria	1	1	1	1	2
Bosnia and Herzegovina	1	2	2	1	2
Bulgaria	1	2	1	0	2
Cyprus	1	1	0	0	1
Croatia	1	1	1	0	2
Egypt	1	1	1	1	1
Spain	1	2	2	1	2
France	1	2	2	1	2
Greece	1	2	1	1	2
Israel	1	1	1	0	1
Italy	1	2	2	1	2
Jordan	1	1	1	0	2
Kosovo	1	1	1	0	2
Lebanon	1	1	1	0	2
Libya	1	1	1	1	1
Macedonia	1	2	1	0	2
Malta	0	0	0	0	0
Morocco	1	2	2	1	2
Montenegro	1	1	0	0	1
Palestine	1	2	2	0	2
Portugal	1	2	2	1	2
Serbia	1	2	1	0	2
Slovenia	1	2	1	0	2
Syria	1	1	1	1	2
Tunisia	1	2	2	1	2
Turkey	1	2	2	2	2

AUTHORS: Doignon Y., Oliveau S. SOURCES: Gadm v2, National Central Statistics Office

Territorial grid 1: First sub-national level in each country
Territorial grid 2: Second subnational level if there are population data, or upper level
Territorial grid 3: Territorial grids harmonized to 500,000 inhabitants
Territorial grid 4: Territorial grids harmonized to 25,000 km²
Territorial grid 5: Second subnational level in each country

ISSN 2282-472X (online)
ISSN 0044-9733 (print)

TABLE 2 – Statistic variability

	Population			Area			Population Density		
	Mean	Coefficient of Variation	Median	Mean	Coefficient of Variation	Median	Mean	Coefficient of Variation	Median
Territorial grid 1	1,194,614	1.72	367,455	21,568	2.72	4308	329	4.34	82
Territorial grid 2	320,233	2.25	66,376	5,856	5.19	764.87	297	4.72	64
Territorial grid 3	626,191	1.46	361,242	11,404	3.66	4234	381	4.02	86
Territorial grid 4	1,521,754	1.24	760,785	27,621	2.31	10,850	293	3.89	79
Territorial grid 5	X	X	X	2,405	6.07	187	X	X	X

AUTHORS: Doignon Y., Oliveau S. SOURCES: Gadm v2, National Central Statistics Office

TABLE 3 – Spatial variability measured by Moran's I, by order of contiguity

	Moran's I (by order of contiguity)				
	Order 1	Order 2	Order 3	Order 4	Order 5
Territorial Grid 1	0.49	0.36	0.24	0.14	0.09
Territorial Grid 2	0.47	0.34	0.24	0.18	0.14
Territorial Grid 3	0.56	0.39	0.29	0.26	0.16
Territorial Grid 4	0.51	0.39	0.25	0.14	0.07

AUTHORS: Doignon Y., Oliveau S. SOURCES: Gadm v2, National Central Statistics Office

ISSN 2282-472X (online)
ISSN 0044-9733 (print)

TABLE 4 – Summary table based on the population criterion

Country	Level 0	Subnational level 1				Subnational level 2			
		Mean	Standard Deviation	Coefficient of Variation	Median	Mean	Standard Deviation	Coefficient of Variation	Median
Albania	3,194,417	266,201	192,204.50	0.72	228,875	89,965	124,649	1.40	48,794
Algeria	29,100,863	606,267	407,915.28	0.67	565,513				
Bosnia and Herzegovina	3,866,530	1,933,265	558,853.36	0.29	X	233,107	173,213	0.74	239,920
Bulgaria	7,563,710	270,132	232,264.15	0.86	198,267	28,695	85,301	2.97	12,267
Cyprus	840,407	140,705	99,141.00	0.70	143,192				
Croatia	4,437,460	211,307	162,453.40	0.77	162,045				
Egypt	68,046,408	2,617,169	1,903,836.50	0.73	2,810,945				
Spain	52,921,889	2,725,559	2,526,866	0.93	1,733,015	860,561	1,148,999	1.34	549,925
France	62,482,931	2,854,041	2,472,616.10	0.87	2,129,309	650,769	494,209	0.76	531,955
Greece	11,260,402	866,184	1,065,867.65	1.23	592,017	220,792	575,323	2.61	113,285
Israel	7,255,300	1,036,471	535,363.46	0.52	1,084,200				
Italy	59,836,894	2,991,844	2,509,262.24	0.84	1,840,867	566,907	647,875	1.14	376,393
Jordan	5,103,639	425,303	551,714.15	1.30	178,894				
Kosovo	1,956,189	391,237	175,467.76	0.45	376,085				
Lebanon	3,755,033	625,838	464,818.42	0.74	436,203				
Libya	6,097,500	190,546	189,484.26	0.99	131,690				
Macedonia	2,048,619	256,077	147,161.52	0.57	201,082	21,276	25,435	1.20	11,928
Malta	416,000								
Morocco	29,607,002	1,973,800	909,417.11	0.46	1,908,905	547,405	463,832	0.85	484,895
Montenegro	628,631	29,934	36,746.13	1.23	18,482				
Palestine	3,767,126	1,883,563	660,466.02	0.35	1,883,563	235,445	149,354	0.63	231,077
Portugal	10,225,836	1,460,833	1,387,922.95	0.95	750,800	354,578	405,062	1.14	248,667
Serbia	7,498,001	299,920	284,715.00	0.95	227,435	46,571	48,052	1.03	27,513
Slovenia	1,990,272	165,856	131,909.12	0.80	122,453	9943	21,722	2.18	4953
Syria	17,920,844	1,280,060	968,152.88	0.76	1 131 587				
Tunisia	9,910,872	1,651,813	1,117,030.58	0.68	1,283,936	412,953	214,180	0.52	403,892
Turkey	67,817,797	10,263,142	5,279,253.12	0.51	9,000,000	941,913	1,606,058	1.71	546,503

AUTHORS: Doignon Y., Oliveau S. SOURCES: National Central Office Statistics

ISSN 2282-472X (online)
ISSN 0044-9733 (print)

TABLE 5 – Summary table based on the surface area criterion

Country	Level 0	Subnational level 1				Subnational level 2				Subnational level 3			
		Mean	Standard Deviation	Coefficient of Variation	Median	Mean	Standard Deviation	Coefficient of Variation	Median	Mean	Standard Deviation	Coefficient of Variation	Median
Albania	28,754	2392	867.92	0.36	2507.78	774	451.78	0.58	809.54	75	54.88	0.72	63.45
Algeria	2,311,292	48,158	114,926.97	2.39	5965.60	1536	8798.88	5.73	121.51				
Bosnia and Herzegovina	50,965	25,483	2573.61	0.10	25,483.49	2831	2192.91541	0.77	2547.00	358	277.72	0.77	291.60
Bulgaria	112,033	4001	1528.52	0.38	3643.17	425	254.43	0.60	371.22				
Cyprus	9269	1544	735.07	0.48	1370.21								
Croatia	56,953	2711	1425.40	0.53	2370.73	101	113.33	1.11	66.95				
Egypt	985,513	37,904	95,282.48	2.51	5160.08								
Spain	505,519	28,083	30,347.28	1.08	10,947.74	9912	5011.16	0.51	9837.47	1373	996.70	0.73	1219.83
France	548,080	24,912	11,434.65	0.46	25,881.15	5709	1939.35	0.34	5929.14	1566	916.00	0.58	1515.25
Greece	132,376	9454	5768.33	0.61	9216.68	2545	1288.09	0.51	2468.70				
Israel	22,420	3202	5104.85	1.59	1240.20								
Italy	300,225	1, 011	7,384.26	0.49	14,340.05	2729	1586.79	0.58	2454.46	37	49.48	1.34	21.67
Jordan	89,190	7432	10,751.68	1.45	2987.30	1715	5331.71	3.11	385.35				
Kosovo	10,887	2180	613.89	0.28	2067.04	363	166.71	0.46	343.86				
Lebanon	10,423	1737	1450.12	0.83	1569.00	400	419.41	1.05	289.14				
Libya	1,618,639	68,020	119,188.12	1.75	13,880.17								
Macedonia	24,791	3098	1005.63	0.32	2974.39	288	232,24	0.81	228.17				
Malta	325												
Morocco	413,764	27,585	24,083.36	0.87	16,554.41	7662	9,081.15	1.19	5245.00	1037	2284.80	2.20	46.00
Montenegro	13,313	655	485.77	0.74	492.40								
Palestine	6225	3112	3,916.66	1.26	3112.37	389	319.31	0.82	317.48				
Portugal	92,140	13,162	13,494.69	1.03	4929.74	3071	2193.50	0.71	2405.53				
Serbia	77,761	3110	974.00	0.31	3046.81	482	284.39	0.59	420.41				
Slovenia	19,932	1660	781.15	0.47	1782.42	103	94.26	0.91	71.16				
Syria	186,521	13,321	13,471.87	1.01	8360.94	3108	5677.66	1.83	1278.96				
Tunisia	154,910	25,817	16,940.39	0.66	19,174.68	6454	8336.76	1.29	4 102.18	578	1950.85	3.38	293.26
Turkey	779,915	111,463	41,692.59	0.37	91,562.15	10,686	6462.88	0.60	8814.78	840	696.67	0.83	649.38

AUTHORS: Doignon Y., Oliveau S. SOURCES: GADM v2

ISSN 2282-472X (online)
ISSN 0044-9733 (print)

Appendix 1
Explanatory insert on the data used in article

Insert on the data

In order to form a grid that is harmonized according to the criteria of population or surface area, the ideal is to have this data for all the administrative levels of the group of countries. The data availability makes it difficult to obtain all of this data. We are focussing exclusively on administrative grids because the data is produced within this legal framework.
For surface areas, we used GADM v2 maps. This data provider makes available a certain amount of administrative content for all the countries in the world. Obviously, the depth of the administrative levels is uneven depending on the country. All levels are available in France, unlike Egypt where only the *mohafazats* level is available. Using a GIS, we have calculated the area of each unit in each available administrative level.
For population, we collected data from the most recent census or annual results from the central statistics offices of each country. For Syria, the current environment meant we were obliged to collect population data from the Gazetteer.de website. For all countries, we did not go below the second administrative level because we have insufficient data at the lower levels for a number of countries.
Table 6 summarizes the administrative levels for which we have surface area and/or population data. This article will only consider those levels for which both population and surface area data is available. Not to have all data for all levels within a country is a disadvantage, but it is possible to demonstrate that, even for only these administrative levels, the grid has a noteworthy effect on the analysis results.

TABLE 6 – Data available for this article

	Administrative Level			
	0	1	2	3
Albania	Population and area	Population and area	Population and area	Surface only
Algeria	Population and area	Population and area	Surface only	
Bosnia and Herzegovina	Population and area	Population and area	Population and area	Surface only
Bulgaria	Population and area	Population and area	Population and area	
Cyprus	Population and area	Population and area		
Croatia	Population and area	Population and area	Surface only	
Egypt	Population and area	Population and area		
Spain	Population and area	Population and area	Population and area	Surface only
France	Population and area	Population and area	Population and area	Surface only
Greece	Population and area	Population and area	Population and area	
Israel	Population and area	Population and area		
Italy	Population and area	Population and area	Population and area	Surface only
Jordan	Population and area	Population and area	Surface only	
Kosovo	Population and area	Population and area	Surface only	
Lebanon	Population and area	Population and area	Surface only	
Libya	Population and area	Population and area		
Macedonia	Population and area	Population and area	Population and area	
Malta	Population and area	Population and area		
Morocco	Population and area	Population and area	Population and area	Surface only
Montenegro	Population and area	Population and area		
Palestine	Population and area	Population and area	Population and area	
Portugal	Population and area	Population and area	Population and area	
Serbia	Population and area	Population and area	Population and area	
Slovenia	Population and area	Population and area	Population and area	
Syria	Population and area	Population and area	Surface only	
Tunisia	Population and area	Population and area	Population and area	Surface only
Turkey	Population and area	Population and area	Population and area	Surface only

Population and area data available for this administrative level
Surface data only available for this administrative level

AUTHORS: Doignon Y., Oliveau S.

ISSN 2282-472X (online)
ISSN 0044-9733 (print)

References

AMRHEIN C.G., REYNOLDS H. (1996), "Using spatial statistics to assess aggregation effects", *Geographical Systems*, 3, pp.143-158.

ANSELIN, L., SYABRI, I., KHO, Y. (2005), "GeoDa : An Introduction to Spatial Data Analysis", *Geographical Analysis*, Vol. 38, n°1, pp. 5-22.

ARBIA G., BENEDETTI R., ESPA G. (1996), "Effects of the MAUP on image classification", *Geographic Systems*, 3, pp.123-141.

CLIFF A.D., ORD K.J. (1973), *Spatial Autocorrelation*, Pion, Londres.

CLIFF A.D., ORD K.J. (1981), *Spatial processes. Models and applications*, Londres, Pion.

DUMOLARD P. (1998), *Validation d'un découpage territorial*, in "Actes des entretiens J. Cartier 1997 : « les découpages du territoire » ", *INSEE-Methodes*, n°76-77-78, pp.59-69.

FOTHERINGHAM A.S., BRUNSDON C., CHARLTON M.E. (2000), "A bluffer's guide to a solution to the ecological inference problem", *Annals of Association of American Geographers*, 90(3), pp.582-586.

GEHLKE C. E., BIEHL K. (1934), "Certain effects of grouping upon the size of the correlation coefficient in census tract material", *Journal of the American Statistical Association*, 29, 185A, pp.169–170.

GRASLAND C., MADELIN M. (2006), Project ESPON 3.4.3, *The modifiable areal unit problem*, final report.

GUILMOTO C.Z., OLIVEAU S. (2007), *Sex Ratio Imbalances among Children at Micro-Level: China and India Compared*, communication in Population Association of America 2007 Annual Meeting, New York, 29-31 mars, http://paa2007.princeton.edu/download.aspx?submissionId=71096.

HOLT D., STEEL D., TRANMER M. (1996), "Area homogeneity and the modifiable areal unit problem", *Geographical Systems*, 3, pp.181-200.

KNUDSEN D.C. (1987), "Computer-intensive significance - testing procedures", *The Professional Geographer*, Vol. 39, pp. 208-214.

MORAN P.A.P. (1950), "Notes on continuous stochastic phenomena", *Biometrika*, Vol. 37, n°1/2, pp.17-23.

OLIVEAU S. (2010), "Autocorrélation spatiale : leçons du changement d'échelle", *L'Espace Géographique*, n°1, pp. 51-64.

OLIVEAU, S., DOIGNON, Y., GUILMOTO, C.Z. (2013), *Neighborhood effects in demography: measuring scales and patterns*, XXVII IUSSP International Population Conference, Session 19-01: Spatial demography, 26-31 août 2013, Busan (Korea).

OPENSHAW S. (1977), "Algorithm 3 : a procedure to generate pseudo-random aggregations of N zones into M zones, where M is less than N", *Environment and Planning* A, 9, pp. 1423-1428.

OPENSHAW S., TAYLOR J. (1979), *A million or so correlation coefficients: Three experiments on the Modifiable Areal Unit Problem*, in WRIGLEY N. (Ed.), *Statistical applications in the spatial science*, Pion Limited, Londres, pp. 127-144.

OPENSHAW S. (1984), "The Modifiable Areal Unit Problem", *Catmog*, n°38, Norwich, Geo Books.

REYNOLDS, H. (1998), *The modifiable area unit problem: Empirical analysis by statistical simulation*, Ph.D. thesis, Department of Geography, University of Toronto, not published, 92 p.

TATE N.J., ATKINSON P.M. (2001), *Modelling scale in geographical information science*, John Wiley & Sons Ltd.

TOBLER W. (1970), A computer movie simulating urban growth in the Detroit region, "Economic Geography", 46(2), pp. 234-240.

ISSN 2282-472X (online)
ISSN 0044-9733 (print)